Performance bounds for coupled models

Chengfang Ren, Rodrigo Cabral Farias, Pierre-Olivier Amblard, Pierre Comon

Performance bounds for coupled models

Chengfang Ren, Rodrigo Cabral Farias, Pierre-Olivier Amblard, and Pierre Comon, *Fellow, IEEE*
CNRS, Gipsa-Lab, University of Grenoble Alpes,
F-38000 Grenoble, France,
Email: firstname.lastname@gipsa-lab.grenoble-inp.fr

***Abstract*—Two models are called "coupled" when a non empty set of the underlying parameters are related through a differentiable implicit function. The goal is to estimate the parameters of both models by merging all datasets, that is, by processing them jointly. In this context, we show that the parameter estimation accuracy under a general class of dataset distributions always improves when compared to an equivalent uncoupled model. We eventually illustrate our results with the fusion of multiple tensor data.**

I. Introduction

Multimodal data analysis is a subject of increasing interest in various domains [1]–[3]. It generally consists in extracting useful information from a collection of datasets acquired by multiple measurement devices [4]. We can distinguish between two types of datasets: homogeneous and heterogeneous. Homogeneous datasets have the same parametric models, while heterogeneous datasets follow different models where some parameters are shared, as described in Section II. Moreover the shared variables do not need to be equal, but can be linked through a (possibly nonlinear) deterministic or statistical relation. In Section III, we derive the Constrained Cramér Rao bound (CCRB) based on the works of [5]–[7]. These approaches can address non-linear coupling links between shared parameters for possibly complex data.

As an illustration, we apply these results in Sec. IV to a problem of coupled complex Canonical Polyadic (CP) decompositions. Complex CP decompositions are relevant in various domains such as antenna array processing, radar and communications [8]–[10]. Algorithms for coupled CP decompositions have been proposed in [11], [12] and performance bounds for their uncoupled models have been proposed in [8], [13], [14]. For coupled real tensors with linear couplings, performance bounds have been proposed in [15], in a Bayesian setting. The present contribution is more general than [15] in some way, since it is not restricted to (real) tensor models and to linear couplings.

II. Coupled datasets model

Let us consider a general coupled model for two datasets $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$:

$$\begin{cases} \boldsymbol{x}_1 \sim f_{\boldsymbol{x}_1;\boldsymbol{\theta}_1,\boldsymbol{\phi}_1} \text{ and } \boldsymbol{x}_2 \sim f_{\boldsymbol{x}_2;\boldsymbol{\theta}_2,\boldsymbol{\phi}_2}, \\ \boldsymbol{g}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \mathbf{0}, \end{cases} \tag{1}$$

where $\boldsymbol{x}_1 \in \Omega_1 \subseteq \mathbb{C}^{n_1}$ and $\boldsymbol{x}_2 \in \Omega_2 \subseteq \mathbb{C}^{n_2}$ are random datasets distributed according to the probability density functions (PDF) $f_{\boldsymbol{x}_1;\boldsymbol{\theta}_1,\boldsymbol{\phi}_1}$ and $f_{\boldsymbol{x}_2;\boldsymbol{\theta}_2,\boldsymbol{\phi}_2}$, respectively parameterized by the unknown deterministic real parameter vectors $(\boldsymbol{\theta}_1,\boldsymbol{\phi}_1) \in \Theta_1 \times \Phi_1 \subseteq \mathbb{R}^{m_1} \times \mathbb{R}^{p_1}$ and $(\boldsymbol{\theta}_2,\boldsymbol{\phi}_2) \in \Theta_2 \times \Phi_2 \subseteq \mathbb{R}^{m_2} \times \mathbb{R}^{p_2}$ (n_i, m_i and p_i denote respectively the size of vectors $\boldsymbol{x}_i$, $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$). The model is referred to as *partially coupled* because only parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ may be linked through the relationship $\boldsymbol{g}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \mathbf{0}$ whereas parameters $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are unrelated. We assume (i) that $\boldsymbol{g}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ is a known *non redundant*[1] deterministic vector function, differentiable for all $(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \in \Theta_1 \times \Theta_2$, (ii) that distribution $f_{\boldsymbol{x}_{1:2};\boldsymbol{\theta}_{1:2},\boldsymbol{\phi}_{1:2}}$ is differentiable w.r.t. $\boldsymbol{\theta}_{1:2}$ and $\boldsymbol{\phi}_{1:2}$ and its support does not depend on these parameters[2], (iii) that variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are statistically independent, *i.e.* $f_{\boldsymbol{x}_{1:2};\boldsymbol{\theta}_{1:2},\boldsymbol{\phi}_{1:2}} = f_{\boldsymbol{x}_1;\boldsymbol{\theta}_1,\boldsymbol{\phi}_1} f_{\boldsymbol{x}_2;\boldsymbol{\theta}_2,\boldsymbol{\phi}_2}$. Moreover, the distributions of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ can be totally different even if the parameter vectors are the same *i.e.* $f_{\boldsymbol{x}_1;\boldsymbol{\theta},\boldsymbol{\phi}} \neq f_{\boldsymbol{x}_2;\boldsymbol{\theta},\boldsymbol{\phi}}$. This scenario corresponds to a data fusion system which collects data from multiple measurement devices.

Under some conditions [5, Lemma 2], the maximum likelihood estimator (MLE) is asymptotically optimal in the mean squared error (MSE) sense, and achieves the CCRB. Here, the term asymptotically means either that the signal to noise ratio (SNR) is high and/or that the ratio between the sizes of the parameter vector and the dataset goes to zero. The CCRB is an insightful substitute to the MSE, especially when the latter is difficult to compute analytically. For these reasons, we subsequently analyze the CCRB for the estimation of $(\boldsymbol{\theta}_{1:2},\boldsymbol{\phi}_{1:2})$ in order to obtain insights on the optimal data fusion performance.

III. Performances analysis

When the coupled model (1) and the associated uncoupled model (*i.e.* the same model without the constraint $\boldsymbol{g}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \mathbf{0}$), are both identifiable, by using the results from [5], the CCRB is given by

$$\mathbf{CCRB} = \boldsymbol{F}^{-1} - \boldsymbol{F}^{-1}\boldsymbol{G}^{\mathsf{T}}\left(\boldsymbol{G}\boldsymbol{F}\boldsymbol{G}^{\mathsf{T}}\right)^{-1}\boldsymbol{G}\boldsymbol{F}^{-1}, \tag{2}$$

where $\boldsymbol{G}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \left[\frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1^{\mathsf{T}}}\ \mathbf{0}\ \frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2^{\mathsf{T}}}\ \mathbf{0}\right]$ is the derivative of $\boldsymbol{g}$ w.r.t. $(\boldsymbol{\theta}_1,\boldsymbol{\phi}_1,\boldsymbol{\theta}_2,\boldsymbol{\phi}_2)$ and matrix $\boldsymbol{F}$ is the Fisher information matrix (FIM) for the uncoupled model. For the sake of simplicity and without loss of generality, we omit in (2) and in what follows the dependency on the true value of parameters $\boldsymbol{F} \triangleq \boldsymbol{F}(\boldsymbol{\theta}_{1:2},\boldsymbol{\phi}_{1:2})$, $\boldsymbol{G} \triangleq \boldsymbol{G}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ and $\mathbf{CCRB} \triangleq$

[1] By non redundant, it is meant that no equation $g_i = 0$ can be deduced from the others, namely $\{g_k = 0,\, k \neq i\}$.

[2] $\boldsymbol{x}_{1:2}$ (resp. $\boldsymbol{\theta}_{1:2}$, $\boldsymbol{\phi}_{1:2}$) denotes the concatenation of vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (resp. $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$).

This work was funded by the FP7 European Research Council Programme, DECODA project, under grant ERC-AdG-2013-320594.

$\mathbf{CCRB}\left(\boldsymbol{\theta}_{1:2}, \boldsymbol{\phi}_{1:2}\right)$. Note that an alternative expression of the general $\mathbf{CCRB}$ is given in [6] when $\boldsymbol{F}$ is singular but the coupled datasets model is still identifiable. Since we want to compare the gain between coupled and uncoupled models, we only focus on the case where they are both identifiable. So, in our case, $\boldsymbol{F}^{-1} = \mathbf{CRB}$. The difference between estimation accuracies for coupled and uncoupled models is given by

$$\mathbf{CRB} - \mathbf{CCRB} = \boldsymbol{F}^{-1}\boldsymbol{G}^{\mathsf{T}}\left(\boldsymbol{G}\boldsymbol{F}^{-1}\boldsymbol{G}^{\mathsf{T}}\right)^{-1}\boldsymbol{G}\boldsymbol{F}^{-1} \succeq \mathbf{0}, \tag{3}$$

where $\boldsymbol{A} \succeq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is a positive semidefinite (PSD) matrix. The inequality (3) means that the CCRB is always lower than the CRB. Coupled models bear additional information, which generally reduce the size of the parameter space and improve estimation performance. In our case, not all parameters are related but only (possibly) $\boldsymbol{\theta}_{1:2}$. So our goal is to study the effect of these partial constraints on the coupled parameters $\boldsymbol{\theta}_{1:2}$ and uncoupled $\boldsymbol{\phi}_{1:2}$ estimation accuracies. In Section III-C, a formula of the CCRB is provided for complex Gaussian models, which can be seen as a modification of the Slepian-Bangs formula [16] for coupled models.

A. Influence on the parameter estimation accuracy

In most performance analyses, we are only interested in the diagonal terms of the CRB, which are directly related to the optimal MSE. Since $f_{\boldsymbol{x}_{1:2};\boldsymbol{\theta}_{1:2},\boldsymbol{\phi}_{1:2}} = f_{\boldsymbol{x}_1;\boldsymbol{\theta}_1,\boldsymbol{\phi}_1} f_{\boldsymbol{x}_2;\boldsymbol{\theta}_2,\boldsymbol{\phi}_2}$ and the pair of variables $(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1)$ and $(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2)$ are independent for the uncoupled model, the classical FIM is a block diagonal matrix $\boldsymbol{F} = \mathrm{Diag}\{\boldsymbol{F}_1, \boldsymbol{F}_2\}$, with

$$\boldsymbol{F}_i = \begin{bmatrix} \partial_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i}\mathcal{L}_i & \partial_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}\mathcal{L}_i \\ \partial_{\boldsymbol{\phi}_i\boldsymbol{\theta}_i}\mathcal{L}_i & \partial_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}\mathcal{L}_i \end{bmatrix}, \text{ for } i = 1, 2, \tag{4}$$

where $\mathcal{L}_i = \ln f_{\boldsymbol{x}_i;\boldsymbol{\theta}_i,\boldsymbol{\phi}_i}(\boldsymbol{x}_i)$ is the log-likelihood function for $\boldsymbol{x}_i$, and $\partial_{\boldsymbol{\theta}_i\boldsymbol{\theta}_j} \triangleq \frac{\partial^2}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j^{\mathsf{T}}}$ is the second derivative operator w.r.t $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j^{\mathsf{T}}$. Due to this special structure, an inversion by blocks of $\boldsymbol{F}$ leads to the following expression:

$$\boldsymbol{F}^{-1} = \mathrm{Diag}\{\mathbf{CRB}_{\boldsymbol{\theta}_1\boldsymbol{\phi}_1}, \mathbf{CRB}_{\boldsymbol{\theta}_2\boldsymbol{\phi}_2}\}. \tag{5}$$

From the above expression, the estimation performances of each pair of variables $(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1)$ and $(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2)$ are independent. Denote $\mathbf{CRB}_{\boldsymbol{\theta}_i}$ and $\mathbf{CRB}_{\boldsymbol{\phi}_i}$ the diagonal blocks of matrix $\mathbf{CRB}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}$ which are respectively the lower bound on the MSE for the estimation of $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ in the uncoupled case. Then

$$\mathbf{CRB}_{\boldsymbol{\theta}_i} = \left(\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i} - \boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}^{-1}\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}^{\mathsf{T}}\right)^{-1}, \tag{6}$$

$$\mathbf{CRB}_{\boldsymbol{\phi}_i} = \left(\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i} - \boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}^{\mathsf{T}}\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i}^{-1}\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}\right)^{-1}, \tag{7}$$

with $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i} = -\mathbb{E}\left[\partial_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i}\mathcal{L}_i\right]$, $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i} = -\mathbb{E}\left[\partial_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}\mathcal{L}_i\right]$ and $\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i} = -\mathbb{E}\left[\partial_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}\mathcal{L}_i\right]$.

In order to take into account the coupling, we need to evaluate the $\mathbf{CCRB}$. Due to the partial coupling $\boldsymbol{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}$, the term $\mathbf{CRB}_{\boldsymbol{\phi}_i}$ vanishes in both products $\boldsymbol{G}\boldsymbol{F}^{-1}$ and $\boldsymbol{F}^{-1}\boldsymbol{G}^{\mathsf{T}}$. The diagonal terms of (3) are usually the most interesting, since they give the gains on MSE for the estimation of $\boldsymbol{\theta}_{1:2}$ and $\boldsymbol{\phi}_{1:2}$. Let us denote $\boldsymbol{K}_{\boldsymbol{\theta}_1}$, $\boldsymbol{K}_{\boldsymbol{\theta}_2}$, $\boldsymbol{K}_{\boldsymbol{\phi}_1}$ and $\boldsymbol{K}_{\boldsymbol{\phi}_2}$ the diagonal blocks of (3) corresponding respectively to the difference between CRB and CCRB for $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. Then by substituting (5) in (3), one has for $i = 1, 2$

$$\boldsymbol{K}_{\boldsymbol{\theta}_i} = \mathbf{CRB}_{\boldsymbol{\theta}_i}\frac{\partial\boldsymbol{g}^{\mathsf{T}}}{\partial\boldsymbol{\theta}_i}\mathbf{CRB}_g^{-1}\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\theta}_i^{\mathsf{T}}}\mathbf{CRB}_{\boldsymbol{\theta}_i} \tag{8}$$

$$\boldsymbol{K}_{\boldsymbol{\phi}_i} = \boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}^{-1}\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}^{\mathsf{T}}\boldsymbol{K}_{\boldsymbol{\theta}_i}\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}^{-1}, \tag{9}$$

where $\mathbf{CRB}_g = \sum_{i=1}^{2}\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\theta}_i^{\mathsf{T}}}\mathbf{CRB}_{\boldsymbol{\theta}_i}\frac{\partial\boldsymbol{g}^{\mathsf{T}}}{\partial\boldsymbol{\theta}_i}$. From these expressions, one can easily check that $\boldsymbol{K}_{\boldsymbol{\theta}_i}$ and $\boldsymbol{K}_{\boldsymbol{\phi}_i}$ are both PSD matrices. As we can see by (9), the presence of a coupling effect between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ also improves the estimation accuracy of $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, despite the absence of a direct dependence.

B. Extremes cases

Note that if there is no link between parameters $\boldsymbol{\theta}_i$ *i.e.* $\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\theta}_i^{\mathsf{T}}} = \mathbf{0}$, then $\boldsymbol{K}_{\boldsymbol{\theta}_i} = \mathbf{0}$ and $\boldsymbol{K}_{\boldsymbol{\phi}_i} = \mathbf{0}$. Consequently, the CCRB is equal to the CRB for the parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$.

On the other hand, if the function $\boldsymbol{g}$ depends only on $\boldsymbol{\theta}_i$ and the derivative $\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\theta}_i^{\mathsf{T}}}$ is a square invertible matrix which means the number of non redundant constraints is equal to the size of the vector $\boldsymbol{\theta}_i$ then $\boldsymbol{K}_{\boldsymbol{\theta}_i} = \mathbf{CRB}_{\boldsymbol{\theta}_i}$ and $\boldsymbol{K}_{\boldsymbol{\phi}_i} = \boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}^{-1}\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}^{\mathsf{T}}\mathbf{CRB}_{\boldsymbol{\theta}_i}\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}^{-1}$. In this case, if we denote $\mathbf{CCRB}_{\boldsymbol{\theta}_i}$ and $\mathbf{CCRB}_{\boldsymbol{\phi}_i}$ the diagonal blocks of $\mathbf{CCRB}$ in (2) for $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ respectively, then one has $\mathbf{CCRB}_{\boldsymbol{\theta}_i} = \mathbf{0}$ and $\mathbf{CCRB}_{\boldsymbol{\phi}_i} = \boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}^{-1}$. This means that $\boldsymbol{\theta}_i$ can be trivially obtained from $\boldsymbol{g}(\boldsymbol{\theta}_i) = \mathbf{0}$ and the vector $\boldsymbol{\phi}_i$ can be estimated independently with the dataset $\boldsymbol{x}_i$ only.

C. Coupled Gaussian observation parameterized by the mean

Let us consider two random complex Gaussian distributed datasets $\boldsymbol{x}_1 \sim \mathbb{C}\mathcal{N}(\mathbf{m}_1(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1), \boldsymbol{\Sigma}_1)$ and $\boldsymbol{x}_2 \sim \mathbb{C}\mathcal{N}(\mathbf{m}_2(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2), \boldsymbol{\Sigma}_2)$ where the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are known and the parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are unknown real and assumed to be deterministic. The functions $\mathbf{m}_1(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1)$ and $\mathbf{m}_2(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2)$ are differentiable w.r.t. $\boldsymbol{\theta}_1, \boldsymbol{\phi}_1$ and $\boldsymbol{\theta}_2, \boldsymbol{\phi}_2$ respectively. In this case, one can show that $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i} = 2\Re\left\{\frac{\partial\mathbf{m}_i^{\mathsf{H}}}{\partial\boldsymbol{\theta}_i}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\mathbf{m}_i}{\partial\boldsymbol{\theta}_i^{\mathsf{T}}}\right\}$, $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i} = 2\Re\left\{\frac{\partial\mathbf{m}_i^{\mathsf{H}}}{\partial\boldsymbol{\theta}_i}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\mathbf{m}_i}{\partial\boldsymbol{\phi}_i^{\mathsf{T}}}\right\}$ and $\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i} = 2\Re\left\{\frac{\partial\mathbf{m}_i^{\mathsf{H}}}{\partial\boldsymbol{\phi}_i}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\mathbf{m}_i}{\partial\boldsymbol{\phi}_i^{\mathsf{T}}}\right\}$, where $\Re$ denotes the real part operator. Then a closed form expression is obtained for $\boldsymbol{K}_{\boldsymbol{\theta}_i}$ and $\boldsymbol{K}_{\boldsymbol{\phi}_i}$ by substituting these three expressions in (8) and (9). These gains correspond to the required modifications of the well-known Slepian-Bangs formula [16] by taking account the coupling link $\boldsymbol{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}$.

IV. Coupled tensor decompositions

In this section, we consider the example of two coupled tensors of order three, $\boldsymbol{\mathcal{X}}^{(1)} \in \mathbb{C}^{I_1\times J_1\times K_1}$ and $\boldsymbol{\mathcal{X}}^{(2)} \in \mathbb{C}^{I_2\times J_2\times K_2}$. Assume these tensors can be modelled by

$$\boldsymbol{\mathcal{X}}^{(1)} = \boldsymbol{\mathcal{T}}^{(1)} + \boldsymbol{\mathcal{N}}^{(1)} \quad \text{and} \quad \boldsymbol{\mathcal{X}}^{(2)} = \boldsymbol{\mathcal{T}}^{(2)} + \boldsymbol{\mathcal{N}}^{(2)}, \tag{10}$$

where $\boldsymbol{\mathcal{T}}^{(1)}$ and $\boldsymbol{\mathcal{T}}^{(2)}$ are two low rank tensors, of rank R_1 and R_2 respectively, and $\boldsymbol{\mathcal{N}}^{(1)}$ and $\boldsymbol{\mathcal{N}}^{(2)}$ are noise tensors. The entries of $\boldsymbol{\mathcal{N}}^{(1)}$ and $\boldsymbol{\mathcal{N}}^{(2)}$ are assumed to be independent and identically distributed (i.i.d.) complex circular Gaussian variables with zero mean and variance σ_1^2 and σ_2^2 respectively, they are also independent from one dataset to another. Let the CP decomposition of tensors $\boldsymbol{\mathcal{T}}^{(i)}$, as defined in [17], be written as follows for $i = 1, 2$

$$\boldsymbol{\mathcal{T}}^{(i)} = \sum_{r=1}^{R_i} \boldsymbol{a}_r^{(i)} \otimes \boldsymbol{b}_r^{(i)} \otimes \boldsymbol{c}_r^{(i)} \text{ subject to } \boldsymbol{g}\left(\boldsymbol{C}^{(1)}, \boldsymbol{C}^{(2)}\right) = \boldsymbol{0}, \quad (11)$$

where operator $\otimes$ denotes the tensor product, and the matrices $\boldsymbol{C}^{(i)} \triangleq \left[\boldsymbol{c}_1^{(i)}\ \boldsymbol{c}_2^{(i)}\ \cdots\ \boldsymbol{c}_{R_i}^{(i)}\right] \in \mathbb{C}^{K_i \times R_i}$. Also define factor matrices $\boldsymbol{A}^{(i)} \triangleq \left[\boldsymbol{a}_1^{(i)}\ \cdots\ \boldsymbol{a}_R^{(i)}\right] \in \mathbb{C}^{I_i \times R_i}$ and $\boldsymbol{B}^{(i)} \triangleq \left[\boldsymbol{b}_1^{(i)}\ \cdots\ \boldsymbol{b}_R^{(i)}\right] \in \mathbb{C}^{J_i \times R_i}$. Since $\boldsymbol{\mathcal{N}}^{(1)}$ and $\boldsymbol{\mathcal{N}}^{(2)}$ are i.i.d., each tensor dataset $\boldsymbol{\mathcal{X}}^{(i)}$ is distributed according to

$$f_{\boldsymbol{\mathcal{X}}^{(i)};\boldsymbol{\theta}_i,\boldsymbol{\phi}_i} = (\pi\sigma_i)^{-I_i J_i K_i} \exp\left(-\frac{1}{\sigma_i^2}\left\|\boldsymbol{\mathcal{X}}^{(i)} - \boldsymbol{\mathcal{T}}^{(i)}\right\|^2\right), \quad (12)$$

where $\|\cdot\|$ is the Frobenius norm: $\|\boldsymbol{\mathcal{X}}\|^2 \triangleq \sum_{i,j,k} |x_{ijk}|^2$.

A. Closed form expression of the CCRB

A tricky vectorization of $\boldsymbol{\mathcal{X}}^{(i)}$ allows to obtain a compact form of the CCRB using the approach presented in Section III-C. To do that, we resort to unfolding matrices. Denote by $\boldsymbol{T}_p^{(i)}$ (resp. $\boldsymbol{X}_p^{(i)}$) the mode-p unfolding[3] matrix of $\boldsymbol{\mathcal{T}}^{(i)}$ (resp. $\boldsymbol{\mathcal{X}}^{(i)}$), $p \in \{1, 2, 3\}$. Then the following relations hold:

$$\begin{aligned} \boldsymbol{T}_1^{(i)} &= \boldsymbol{A}^{(i)}(\boldsymbol{C}^{(i)} \odot \boldsymbol{B}^{(i)})^{\mathsf{T}}, \\ \boldsymbol{T}_2^{(i)} &= \boldsymbol{B}^{(i)}(\boldsymbol{C}^{(i)} \odot \boldsymbol{A}^{(i)})^{\mathsf{T}}, \\ \boldsymbol{T}_3^{(i)} &= \boldsymbol{C}^{(i)}(\boldsymbol{B}^{(i)} \odot \boldsymbol{A}^{(i)})^{\mathsf{T}}, \end{aligned} \quad (13)$$

where operator $\odot$ denotes the Khatri-Rao product [19]. Consequently, the vectorization of $\boldsymbol{T}_3^{(i)}$ leads to [19]:

$$\operatorname{vec} \boldsymbol{T}_3^{(i)} = \boldsymbol{S}_C \operatorname{vec} \boldsymbol{C}^{(i)} = \boldsymbol{S}_A \operatorname{vec} \boldsymbol{A}^{(i)} = \boldsymbol{S}_B \operatorname{vec} \boldsymbol{B}^{(i)} \quad (14)$$

where $\boldsymbol{S}_C = \boldsymbol{B}^{(i)} \odot \boldsymbol{A}^{(i)} \boxtimes \boldsymbol{I}_{K_i}$, $\boldsymbol{S}_A = \boldsymbol{J}_{13}^{(i)}(\boldsymbol{C}^{(i)} \odot \boldsymbol{B}^{(i)} \boxtimes \boldsymbol{I}_{I_i})$, $\boldsymbol{S}_B = \boldsymbol{J}_{23}^{(i)}(\boldsymbol{C}^{(i)} \odot \boldsymbol{A}^{(i)} \boxtimes \boldsymbol{I}_{J_i})$, $\boldsymbol{J}_{23}^{(i)}$ (resp. $\boldsymbol{J}_{13}^{(i)}$) is the permutation matrix mapping the entries of $\operatorname{vec} \boldsymbol{X}_2^{(i)}$ to those of $\operatorname{vec} \boldsymbol{X}_3^{(i)}$ (resp. $\operatorname{vec} \boldsymbol{X}_1^{(i)}$ to those of $\operatorname{vec} \boldsymbol{X}_3^{(i)}$), operator $\boxtimes$ denotes the Kronecker product, and $\boldsymbol{I}_\alpha$ stands for the identity matrix of size α.

The scenario described in (10) corresponds to a coupled model defined in (1) with $\boldsymbol{x}_i = \operatorname{vec} \boldsymbol{X}_3^{(i)} \in \mathbb{C}^{I_i J_i K_i}$, $\boldsymbol{\theta}_i = [(\Re\ \Im)\{\operatorname{vec}^{\mathsf{T}} \boldsymbol{C}^{(i)}\}]^{\mathsf{T}} \in \mathbb{R}^{2 I_i R_i}$, and $\widetilde{\boldsymbol{\phi}}_i = [(\Re\ \Im)\{\operatorname{vec}^{\mathsf{T}} \boldsymbol{A}^{(i)}\}\ (\Re\ \Im)\{\operatorname{vec}^{\mathsf{T}} \boldsymbol{B}^{(i)}\}]^{\mathsf{T}} \in \mathbb{R}^{2(J_i+K_i)R_i}$. Clearly with the above notations, the dataset distribution is

$$\boldsymbol{x}_i \sim \mathbb{C}\mathcal{N}(\boldsymbol{m}_i(\boldsymbol{\theta}_i, \widetilde{\boldsymbol{\phi}}_i), \sigma_i^2 \boldsymbol{I}), \quad (15)$$

where $\boldsymbol{m}_i(\boldsymbol{\theta}_i, \widetilde{\boldsymbol{\phi}}_i) = \operatorname{vec} \mathbf{T}_3^{(i)}$. However model (15) is not identifiable since the CP decomposition (11) is unique only up to a scaling factor [17]. To fix this indeterminacy, a solution is to fix the elements of the first row of the matrices $\boldsymbol{A}^{(i)}$ and $\boldsymbol{B}^{(i)}$ to known values. This in turn specifies some elements in $\widetilde{\boldsymbol{\phi}}_i$, and therefore, we must consider $\boldsymbol{\phi}_i \in \mathbb{R}^{2(J_i+K_i-2)R_i}$ containing only the unknown parameters of $\widetilde{\boldsymbol{\phi}}_i$. The relation between $\boldsymbol{\phi}_i$ and $\widetilde{\boldsymbol{\phi}}_i$ is given by $\boldsymbol{\phi}_i = \boldsymbol{M}\widetilde{\boldsymbol{\phi}}_i$ where $\boldsymbol{M} \in \mathbb{R}^{2(J_i+K_i-2)R_i \times 2(J_i+K_i)R_i}$ is a mask matrix which is obtained from an identity matrix $\boldsymbol{I}_{2(J_i+K_i)R_i}$ by removing the $4R_i$ rows corresponding to the known parameters of $\widetilde{\boldsymbol{\phi}}_i$.

As seen in Sec. III-C, a closed form expression of $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i}$, $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}$ and $\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}$ is desirable in order to compute the CCRB. Using (14), the partial derivatives below can be obtained

$$\frac{\partial \boldsymbol{m}_i(\boldsymbol{\theta}_i, \boldsymbol{\phi}_i)}{\partial \boldsymbol{\theta}_i^{\mathsf{T}}} = \boldsymbol{S}_C \left[\boldsymbol{I}_{K_i R_i}\ j\boldsymbol{I}_{K_i R_i}\right] \quad (16)$$

$$\frac{\partial \boldsymbol{m}_i(\boldsymbol{\theta}_i, \boldsymbol{\phi}_i)}{\partial \boldsymbol{\phi}_i^{\mathsf{T}}} = \left[\boldsymbol{S}_A\ \boldsymbol{S}_B\right] \boldsymbol{P}\boldsymbol{M}^{\mathsf{T}} \quad (17)$$

with $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I}_{I_i R_i} & j\boldsymbol{I}_{I_i R_i} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_i R_i} & j\boldsymbol{I}_{J_i R_i} \end{bmatrix}$, j the pure imaginary unit and $\boldsymbol{I}_{\alpha\beta}$ the identity matrix of size $\alpha\beta$. Using these expressions in $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i}$, $\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i}$ and $\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i}$, one gets

$$\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i} = \frac{2}{\sigma_i^2}\Re\left\{\left[\boldsymbol{I}_{K_i R_i}\ j\boldsymbol{I}_{K_i R_i}\right]^{\mathsf{H}} \boldsymbol{S}_C^{\mathsf{H}} \boldsymbol{S}_C \left[\boldsymbol{I}_{K_i R_i}\ j\boldsymbol{I}_{K_i R_i}\right]\right\} \quad (18)$$

$$\boldsymbol{D}_{\boldsymbol{\theta}_i\boldsymbol{\phi}_i} = \frac{2}{\sigma_i^2}\Re\left\{\left[\boldsymbol{I}_{K_i R_i}\ j\boldsymbol{I}_{K_i R_i}\right]^{\mathsf{H}} \boldsymbol{S}_C^{\mathsf{H}} \left[\boldsymbol{S}_A\ \boldsymbol{S}_B\right] \boldsymbol{P}\boldsymbol{M}^{\mathsf{T}}\right\} \quad (19)$$

$$\boldsymbol{D}_{\boldsymbol{\phi}_i\boldsymbol{\phi}_i} = \frac{2}{\sigma_i^2}\Re\left\{\boldsymbol{M}\boldsymbol{P}^{\mathsf{H}} \begin{bmatrix} \boldsymbol{S}_A^{\mathsf{H}}\boldsymbol{S}_A & \boldsymbol{S}_B^{\mathsf{H}}\boldsymbol{S}_A \\ \boldsymbol{S}_A^{\mathsf{H}}\boldsymbol{S}_B & \boldsymbol{S}_B^{\mathsf{H}}\boldsymbol{S}_B \end{bmatrix} \boldsymbol{P}\boldsymbol{M}^{\mathsf{T}}\right\} \quad (20)$$

with $\boldsymbol{S}_C^{\mathsf{H}}\boldsymbol{S}_C = \left(\boldsymbol{B}^{(i)\mathsf{H}}\boldsymbol{B}^{(i)} \boxdot \boldsymbol{A}^{(i)\mathsf{H}}\boldsymbol{A}^{(i)}\right) \boxtimes \boldsymbol{I}_{K_i}$, where the operator $\boxdot$ denotes the Hadamard (entry-wise) product. Consequently, the CCRB in (2) is obtained by plugging expressions (18), (19), (20) in (5), (8) and (9).

B. Simulations

In this section we simulate the estimation performance of the complex CP model under additive complex circular Gaussian noise and we compare it with the Cramér-Rao bounds given in Subsec. IV-A. We assume an equality constraint on the coupled factors, that is, $\boldsymbol{g}\left(\boldsymbol{C}^{(1)}, \boldsymbol{C}^{(2)}\right) = \boldsymbol{C}^{(1)} - \boldsymbol{C}^{(2)}$. The CP parameters are retrieved using MLE. Since the ML objective for CP decompositions exhibits many local minima, most existing approaches execute repeated descent algorithms initialized at different random points; the execution that led to the largest objective value is eventually selected as estimate [20, pp. 62-63, 122].

Uncoupled ALS: A standard method to obtain stationary points of the negative log-likelihood in the uncoupled case is block coordinate descent, that is, minimization of the objective function w.r.t. one block while the others are fixed to their last update value. From (14), by choosing blocks corresponding to factors $\boldsymbol{A}^{(i)}$, $\boldsymbol{B}^{(i)}$ and $\boldsymbol{C}^{(i)}$, block-wise minimization merely corresponds to a linear least-squares problem, and is known as the Alternating Least Squares (ALS) algorithm [20, p. 63]. For instance, the LS solution for factor $\boldsymbol{C}^{(i)}$ at the k-th iterate can be shown to be

$$\widehat{\boldsymbol{C}}_k^{(i)} = \boldsymbol{X}_3^{(i)} \left(\widehat{\boldsymbol{B}}_k^{(i)} \odot \widehat{\boldsymbol{A}}_k^{(i)}\right)^{\star} \left(\boldsymbol{M}_k^{(i)}\right)^{-1},$$

[3] See *e.g.* [17], [18] for a definition of mode-p unfoldings and properties.

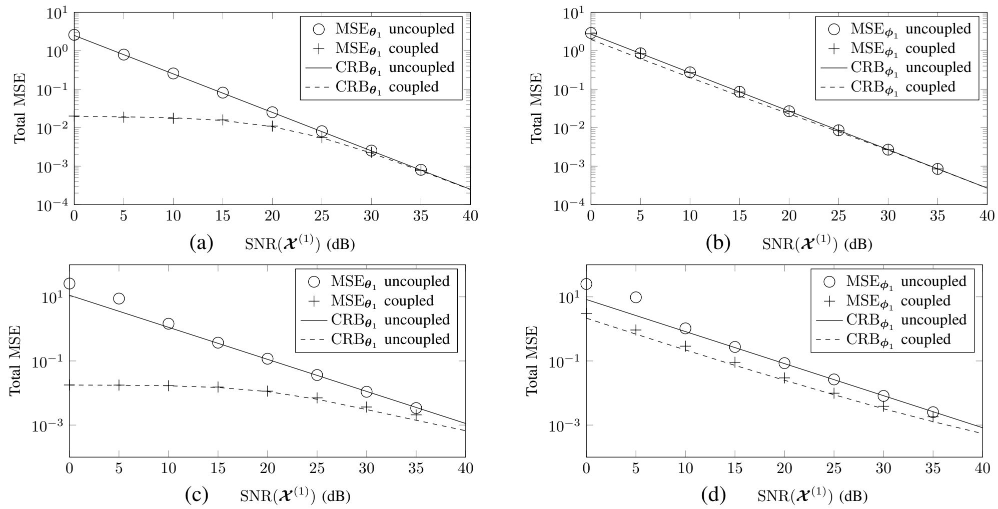

Fig. 1. Total mean squared error and performance bounds for the estimation of complex coupled CP tensor models. Uncoupled parameters $\boldsymbol{\phi}_{1:2}$ are vectorized real and imaginary parts of $\boldsymbol{A}^{(1:2)}$ and $\boldsymbol{B}^{(1:2)}$, while coupled parameters (equality constraint) $\boldsymbol{\theta}_{1:2}$ are vectorized real and imaginary parts of $\boldsymbol{C}^{(1:2)}$. SNR of $\boldsymbol{\mathcal{X}}^{(2)}$ is fixed to 20dB whereas SNR $\left(\boldsymbol{\mathcal{X}}^{(1)}\right) \in [0,\, 40]\,\text{dB}$. Results for i.i.d. distributed $\boldsymbol{A}^{(1)}$ are shown in (a) and (b), and for $\boldsymbol{A}^{(1)}$ with two collinear columns are shown in (c) and (d).

where $(^\star)$ denotes complex conjugation and $\boldsymbol{M}_k^{(i)} = \left(\widehat{\boldsymbol{B}}_k^{(i)\mathsf{T}} \widehat{\boldsymbol{B}}_k^{(i)\star}\right) \boxdot \left(\widehat{\boldsymbol{A}}_k^{(i)\mathsf{T}} \widehat{\boldsymbol{A}}_k^{(i)\star}\right)$. Similar expressions can be obtained for the update of the other factors.

Coupled ALS: in the coupled problem, the ALS update of the uncoupled parameters are not modified while the update of the constrained factors can be done by jointly estimating only one factor $\boldsymbol{C}$:

$$\widehat{\boldsymbol{C}}_k = \left[\sum_{i=1}^{2} \frac{1}{\sigma_i^2} \boldsymbol{X}_3^{(i)} \left(\widehat{\boldsymbol{B}}_k^{(i)} \odot \widehat{\boldsymbol{A}}_k^{(i)}\right)^\star\right] \left[\sum_{i=1}^{2} \frac{1}{\sigma_i^2} \boldsymbol{M}_k^{(i)}\right]^{-1}.$$

To speed up convergence of the coupled ALS, the factors obtained with uncoupled ALS initialized randomly can be used. In this way, the initialization is expected to be closer to the solution than random initialization. Note that to properly initialize coupled algorithms, scaling and permutation ambiguities inherent to CP models need to be corrected. Scaling ambiguity can be corrected by normalizing $\boldsymbol{A}^{(i)}$ and $\boldsymbol{B}^{(i)}$ such that $\hat{a}_{1r}^{(i)} = 1$ and $\hat{b}_{1r}^{(i)} = 1$, whereas permutation ambiguity can be corrected by searching for the best column permutation in $\widehat{\boldsymbol{C}}^{(2)}$ for fixed $\widehat{\boldsymbol{C}}^{(1)}$.

Simulation settings: the tensors we consider have all dimensions equal to 15 and $R^{(1)} = R^{(2)} = 3$. In one set of simulations all CP factors are generated randomly according to i.i.d. standard circular complex Gaussian variables, while in the other all factors are i.i.d. except factor $\boldsymbol{A}^{(1)}$, which is generated with two nearly collinear columns (correlation coefficient $\rho = 0.99$). First rows of the uncoupled factors are set to 1. We evaluate the total MSE on the real and imaginary parts of $\widehat{\boldsymbol{\theta}}_i$ (coupled factor $\boldsymbol{C}^{(i)}$) and $\widehat{\boldsymbol{\phi}}_i$ (uncoupled factors $\boldsymbol{A}^{(i)}$ and $\boldsymbol{B}^{(i)}$) by averaging the squared errors through 500 noise realizations. For each realization 5 different initializations are used. The SNR[4] of $\boldsymbol{\mathcal{X}}^{(2)}$ is fixed to 20dB while that of $\boldsymbol{\mathcal{X}}^{(1)}$ varies from 0 to 40dB. The results for the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\phi}_1$ are shown in Fig. 1a, 1b for i.i.d. $\boldsymbol{A}^{(1)}$ and in Fig. 1c, 1d for $\boldsymbol{A}^{(1)}$ with two nearly collinear columns. In these figures, we have also plotted the performance bounds for uncoupled and coupled models.

We can see in Fig. 1a and 1c that the CRB gains on $\boldsymbol{\theta}_1$ obtained with data fusion can be large in both settings. However, the predicted gains on $\boldsymbol{\phi}_1$ are small for i.i.d. $\boldsymbol{A}^{(1)}$. For the simulated model, we can also observe that total MSE curves follow closely the performance bounds for $\boldsymbol{\theta}_1$. For $\boldsymbol{\phi}_1$, the small predicted estimation gain when $\boldsymbol{A}^{(1)}$ is i.i.d. cannot be observed, whereas for $\boldsymbol{A}^{(1)}$ with nearly collinear columns, total MSE follows the CRB with a small gap. It seems that for CP decompositions, both in theory and in practice, data fusion helps most when the problem is badly conditioned.

V. Conclusions

We have proposed a parametric estimation perspective on data fusion of multimodal datasets. When the inherent dependencies between two datasets are correctly modelled as constraints linking some of their underlying parameters, performance of joint parameter estimation can be lower bounded by constrained Cramér-Rao bounds. With such an approach, we demonstrate quantitatively a fact that was intuitively expected: (asymptotically) optimal data fusion enhances parameter estimation performance, even for parameters that are not directly linked. An illustration related to coupled CP decompositions of complex tensors has been eventually provided.

[4]SNR for complex random tensors is $\text{SNR}(\boldsymbol{\mathcal{X}}^{(i)}) = 10\log_{10}\left\{\frac{\mathbb{E}[||\boldsymbol{\mathcal{T}}^{(i)}||^2]}{\mathbb{E}[||\boldsymbol{\mathcal{X}}^{(i)} - \boldsymbol{\mathcal{T}}^{(i)}||^2]}\right\} = 10\log_{10}\left(\frac{R_i}{\sigma_i^2}\right)$, where σ_i is the variance of the complex Gaussian noise. .

REFERENCES

[1] J. Sui, T. Adalı, Q. Yu, J. Chen, and V. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," *J. Neurosci. Meth.*, vol. 204, no. 1, pp. 68–81, 2012.

[2] E. Acar, A. Lawaetz, M. Rasmussen, and R. Bro, "Structure-revealing data fusion model with applications in metabolomics," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.* IEEE, 2013, pp. 6023–6026.

[3] B. Ermiş, E. Acar, and A. T. Cemgil, "Link prediction in heterogeneous data via generalized coupled tensor factorization," *Data Min. Knowl. Discov.*, vol. 29, no. 1, pp. 203–236, 2013.

[4] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[5] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.

[6] P. Stoica and B. C. Ng, "On the cramér-rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, 1998.

[7] T. J. Moore Jr., "A theory of cramer-rao bounds for constrained parametric models," Ph.D. dissertation, University of Maryland, 2010.

[8] S. Sahnoun and P. Comon, "Joint source estimation and localization," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2485–2495, 2015.

[9] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind parafac receivers for ds-cdma systems," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, 2000.

[10] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for mimo radar," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5693–5705, 2010.

[11] E. Acar, T. G. Kolda, and D. M. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations," *arXiv preprint arXiv:1105.3422*, 2011.

[12] K. Y. Yılmaz, A. T. Cemgil, and U. Simsekli, "Generalised coupled tensor factorisation," in *Adv. Neural Inf. Process. Syst.*, 2011, pp. 2151–2159.

[13] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *Signal Processing, IEEE Transactions on*, vol. 49, no. 9, pp. 2074–2086, 2001.

[14] M. Boizard, R. Boyer, G. Favier, J. E. Cohen, and P. Comon, "Performance estimation for tensor CP decomposition with structured factors," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3482–3486.

[15] R. Cabral Farias, J. E. Cohen, and P. Comon, "Exploring multimodal data fusion through joint decompositions with flexible couplings," *IEEE Trans. Signal Process.*, May 2015, submitted, arxiv:1505.07717.

[16] D. Slepian, "Estimation of signal parameters in the presence of noise," *Trans. IRE Professional Group Inf. Theory*, vol. 3, no. 3, pp. 68–69, 1954.

[17] P. Comon, "Tensors: a brief introduction," *IEEE Signal Process. Magazine*, vol. 31, no. 3, pp. 44–53, May 2014, special issue on BSS. hal-00923279.

[18] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[19] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. on Circuits and Systems*, vol. 25, no. 9, pp. 114–122, Sep. 1978.

[20] R. Bro, "Multi-way analysis in the food industry: Models, algorithms, and applications," Ph.D. dissertation, University of Amsterdam, The Netherlands, 1998.